

### Hinweise zum Datenschutz: Rechtlicher Rahmen und Maßnahmen zur datenschutzgerechten Archivierung sozialwissenschaftlicher Forschungsdaten

Kinder-Kurlanda, Katharina; Watteler, Oliver

Veröffentlichungsversion / Published Version

Monographie / monograph

**Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:**

GESIS - Leibniz-Institut für Sozialwissenschaften

#### Empfohlene Zitierung / Suggested Citation:

Kinder-Kurlanda, K., & Watteler, O. (2015). *Hinweise zum Datenschutz: Rechtlicher Rahmen und Maßnahmen zur datenschutzgerechten Archivierung sozialwissenschaftlicher Forschungsdaten*. (GESIS Papers, 2015/01). Köln: GESIS - Leibniz-Institut für Sozialwissenschaften. <https://doi.org/10.21241/ssoar.43183>

#### Nutzungsbedingungen:

Dieser Text wird unter einer CC BY-NC Lizenz (Namensnennung-Nicht-kommerziell) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:  
<https://creativecommons.org/licenses/by-nc/4.0/deed.de>

#### Terms of use:

This document is made available under a CC BY-NC Licence (Attribution-NonCommercial). For more information see:  
<https://creativecommons.org/licenses/by-nc/4.0>

## Hinweise zum Datenschutz

Rechtlicher Rahmen und Maßnahmen  
zur datenschutzgerechten Archivierung  
sozialwissenschaftlicher Forschungsdaten

*Katharina Kinder-Kurlanda & Oliver Watteler*



GESIS-Papers 2015|01

## Hinweise zum Datenschutz

Rechtlicher Rahmen und Maßnahmen zur datenschutzgerechten Archivierung sozialwissenschaftlicher Forschungsdaten

*Katharina Kinder-Kurlanda & Oliver Watteler*

## GESIS-Papers

GESIS – Leibniz-Institut für Sozialwissenschaften

Datenarchiv für Sozialwissenschaften

Unter Sachsenhausen 6-8

50667 Köln

Telefon: (0221) 476 94 - 449 / -418

Telefax: (0221) 476 94 - 199

E-Mail: [katharina.kinder-kurlanda@gesis.org](mailto:katharina.kinder-kurlanda@gesis.org) / [oliver.watteler@gesis.org](mailto:oliver.watteler@gesis.org)

ISSN: 2364-3781 (Online)

Herausgeber,

Druck und Vertrieb:

GESIS – Leibniz-Institut für Sozialwissenschaften  
Unter Sachsenhausen 6-8, 50667 Köln

# 1 Einleitung

---

In der sozialwissenschaftlichen Community und bei Forschungsförderern wie der Deutschen Forschungsgemeinschaft (DFG) und dem Bundesministerium für Bildung und Forschung (BMBF) gibt es großes Interesse daran, Ergebnisse der Forschung langfristig zu archivieren. Eine Archivierung bietet Dritten Möglichkeiten, publizierte Ergebnisse zu überprüfen, die Daten unter neuen Fragestellungen auszuwerten, neue statistische Verfahren auf die Daten anzuwenden oder die Daten und ihre Erhebungsmethoden zur Basis eigener Datenerhebungen zu machen. Um bei einer Archivierung die Persönlichkeitsrechte der Untersuchungspersonen zu wahren, muss der rechtliche Rahmen der Datenerhebung eine Archivierung zulassen und die Daten müssen in der Regel anonymisiert werden.

Im Folgenden wird das Thema Datenschutz in der sozialwissenschaftlichen Forschung in Bezug auf die Nutzung der Daten durch Dritte behandelt. Haben Wissenschaftler in Forschungsvorhaben die Möglichkeit über informierte Einwilligungen („informed consent“) Vereinbarungen mit Untersuchungspersonen zu schließen, müssen bei einer Weitergabe der Daten an Dritte, nicht am Vorhaben beteiligte Personen datenschutzrechtliche Bestimmungen berücksichtigt werden. Die Anonymisierung der Daten stellt dabei häufig den zentralen Arbeitsschritt auf dem Weg von Daten in Forschungsdatenzentren und Datenarchive dar. Das vorliegende Handout dient in erster Linie als praxisnahe Darstellung von Maßnahmen, wie mit Forschungsdaten umgegangen werden kann, um sie für die Nachnutzung zu publizieren.

Nach einer kurzen Einführung in die rechtliche Situation im Anschluss an die Datenerhebung in Forschungsprojekten werden Maßnahmen diskutiert, die den Schutz von Personen im Sinne des Datenschutzrechtes sicherstellen sollen. Hier wird vor allem auf die Konzepte der Anonymisierung, Pseudonymisierung und der Zugangskontrolle eingegangen. Als Ergänzung zu den informationsverändernden Verfahren der Anonymisierung und Pseudonymisierung werden im Anhang neun Fallbeispiele aus der Praxis datenhaltender Einrichtungen, wie es zum Beispiel GESIS ist, vorgestellt. Den Abschluss bildet eine nicht vollständige Liste möglicher Identifizierungsmerkmale, auf die bei einer Bearbeitung von Daten, die von Individuen erhoben wurden, geachtet werden sollte.

An dieser Stelle sei klargestellt, dass jeder Datenbestand seine eigenen Charakteristika wie etwa Komplexität, Umfang an erhobenen Informationen oder spezielle Erhebungsdesigns hat. D.h., dieses Handout dient in erster Linie als Leitfaden. Für den Datenschutz in der sozialwissenschaftlichen Forschung gibt es keine standardisierten Lösungen.

## 2 Datenschutz in der sozialwissenschaftlichen Forschung

---

### 2.1 Rechtlicher Rahmen

In den empirischen Sozial- und Wirtschaftswissenschaften werden regelmäßig Untersuchungen über Personen angestellt. Der Umgang mit diesen Personen, deren Verhalten, Ansichten, Einstellungen oder soziale und wirtschaftliche Lebensbedingungen analysiert werden, unterliegt besonderen forschungsethischen Rahmenbedingungen. Die sich aus Forschungsvorhaben eventuell ergebenden Folgen für einzelne TeilnehmerInnen sind dabei eine wichtige Kategorie der Forschungsethik. Diese Form der Ethik kann man als Reflexion über Forschung als Praxis verstehen. Zu den möglichen Folgen gehört u.a. die Schädigung von Versuchspersonen, weshalb rechtliche Regelungen und freiwillige Verpflichtungen zum Schutz der Personen allgemein anerkannt sind (Graumann 2006: 253-256).<sup>1</sup> D.h. es müssen bei der Forschung Persönlichkeitsrechte (Palandt 2008: 1216-1226) beachtet werden, wie z.B. das Recht auf informationelle Selbstbestimmung, das Recht am eigenen Bild oder das Recht auf Privat- und Intimsphäre. Ein besonderes Schutzrecht ist der Datenschutz (Bethlehem 2009; Höhne 2010; Metschke, Wellbrock 2003; Wirth 1992 und 2003), also der Schutz des Einzelnen davor, „durch den Umgang mit seinen personenbezogenen Daten in seinem Persönlichkeitsrecht beeinträchtigt“ zu werden (§1, 1 Bundesdatenschutzgesetz – BDSG).

Grundsätzlich muss die Teilnahme an einem Forschungsprojekt im Rahmen eines „informed consent“, also eines informierten Einverständnisses erfolgen.<sup>2</sup> Hierunter versteht man, dass den Personen ausreichende Informationen über den Zweck der Datenerhebung und -verarbeitung, sprich das Forschungsvorhaben zur Verfügung gestellt werden, ihnen die Freiwilligkeit der Teilnahme zugesichert wird und sie auf der Basis der genannten Informationen eine Entscheidung über eine Teilnahme treffen können. Liegen Einwilligungen der Befragten zur Verarbeitung ihrer Daten zu Zwecken der Forschung vor, so ist diese Verarbeitung rechtlich unproblematisch. Man spricht von ausgeübtem informationellem Selbstbestimmungsrecht. Das Prinzip des „informed consent“ haben mehrere wissenschaftliche Berufsverbände, u.a. die Deutsche Gesellschaft für Soziologie (DGS), als für ihre Mitglieder verbindliche Vereinbarung formuliert.<sup>3</sup> Ist eine Datenweitergabe in anonymisierter oder anderer Form in der Einverständniserklärung explizit ausgeschlossen, können die Daten nur durch die PrimärforscherInnen genutzt werden. Eine Archivierung an anderem Ort ist ausgeschlossen.

Die Beurteilung des Datenschutzes für die Archivierung oder das Angebot sozialwissenschaftlicher Daten über ein Forschungsdatenzentrum (FDZ) erfolgt auf zwei Stufen:

Auf einer ersten Stufe betrachtet man den Erhebungskontext der Daten. Wann wurden welche Daten zu welchem Zweck erhoben? Hierbei kann es sich um Daten handeln, die bei Individuen erfragt oder aus Quellen wie Registern zusammengestellt wurden. Wichtig ist der rechtliche Rahmen, in dem die Erhebung stattfand. Mussten z.B. spezielle Gesetze beachtet werden (z.B. Landesschulgesetze), war die Teilnahme vorgeschrieben (z.B. Mikrozensus) oder freiwillig, gab es vertragliche Abmachungen (z.B. bei Befragungen von Unternehmen) etc.

---

<sup>1</sup> Vgl. zur Umsetzung ethischer Rahmenbedingungen durch wissenschaftliche Fachverbände z.B. den Ethik-Kodex der Deutschen Gesellschaft für Soziologie (DGS).

<sup>2</sup> Das Prinzip des „informed consent“ hat seinen historischen Ursprung im sog. Nürnberger Kodex von 1946. Zur historischen Entwicklung des Prinzips und seiner Basis in der Medizin vgl. Faden/Beauchamp (1986).

<sup>3</sup> Vgl. Ethik-Kodex, Abs. B, Satz 3: „Generell gilt für die Beteiligung an sozialwissenschaftlichen Untersuchungen, daß diese freiwillig ist und auf der Grundlage einer möglichst ausführlichen Information über Ziele und Methoden des entsprechenden Forschungsvorhabens erfolgt. [...]“

Nach der Überprüfung der rechtlichen Situation können die Daten dahingehend geprüft werden, ob Variablen in ihrem Informationsgehalt vergrößert werden müssen. Direkte Identifizierungsmerkmale sollten prinzipiell gelöscht werden (formale Anonymisierung; vgl. Anhang 3 für direkte Identifizierungsmerkmale). Alle weiteren Maßnahmen hängen z.B. in allgemeinen Bevölkerungsumfragen in der Regel von der Verortung und den beruflichen Angaben der Befragten ab. D.h., umso detaillierter die regionalen Angaben in den Daten (z.B. Ortsnamen, Gemeindekennziffern) und umso genauer die Nennung von Berufen (z.B. Berufsbezeichnung im Klartext, 4-stelliger ISCO-Code)<sup>4</sup>, desto höher ist das Risiko einer Reidentifizierung.<sup>5</sup> Umso spezieller die Erhebung (regionale Beschränkung, besondere Untersuchungspopulationen wie sog. Eliten, Lebensverlaufsuntersuchungen mit vielen biographischen Details etc.), desto mehr muss auf weitere Merkmalskombinationen geachtet werden. Der Anhang 2 bietet eine Übersicht über Fallbeispiele aus der Archivierungspraxis. Anhang 3 bietet eine nicht vollständige Liste von Merkmalen, auf die man bei einer Überprüfung achten sollte.

## 2.2 Maßnahmen zur Sicherstellung des Datenschutzes

Wenn die Daten in ihrem Informationsgehalt reduziert werden müssen, um eine Reidentifizierung zu vermeiden, spricht man von Anonymisierung und Pseudonymisierung. Die Daten müssen dann in aller Regel der Anforderung der „faktischen Anonymität“ genügen. Maßnahmen zum Schutz der Daten und damit der hinter den Daten stehenden Personen müssen diskutiert werden. Hierzu zählen eventuelle Datenmodifikationen wie die Recodierung extremer Fälle bei einzelnen Angaben (z.B. bei Einkommen und Alter), die Verringerung des Detailgrades (z.B. bei Berufsangaben), oder die Entnahme einzelner Variablen. Weiter kann der Datenzugang eingeschränkt oder eine vertragliche Vereinbarung mit Dritten für eine Datennutzung geschlossen werden.

### 2.2.1 Anonymisierung (z.B. Reduzierung von Variablen)

Anonymisierung bedeutet im Sinne des Bundesdatenschutzgesetz (§3, Abs.6) das „Verändern personenbezogener Daten derart, dass die Einzelangaben über persönliche oder sachliche Verhältnisse nicht mehr oder nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft einer bestimmten oder bestimmaren natürlichen Person zugeordnet werden können.“ D.h. Angaben im Datensatz werden so verändert, dass die Auskunft gebende Person nur mit großem Aufwand bestimmbar ist. Maßnahmen zum Schutz der Daten und damit der hinter den Daten stehenden Personen sind z.B. die Recodierung extremer Fälle bei einzelnen Angaben (z.B. bei sehr hohen Einkommen und hohen Altersangaben), die Verringerung des Detailgrades (z.B. bei Berufsangaben), oder die Entnahme einzelner Variablen. Weiter kann der Datenzugang eingeschränkt oder eine vertragliche Vereinbarung mit Dritten für eine Datennutzung geschlossen werden. Prinzipiell ist die Verwendung anonymisierter Daten z.B. für Forschungszwecke zulässig.

Der durch Anonymisierungsmaßnahmen erreichte Zustand wird auch als „faktische Anonymität“ bezeichnet. Daneben gibt es in der Forschung die Begriffe der formalen Anonymisierung, also der Entfernung direkter Merkmale wie Name, Anschrift, etc. (siehe Anhang 3), sowie der absoluten Anonymisierung, d.h. der Veränderung der Daten dahingehend, dass auch mit Hilfe von Zusatzwissen eine Reidentifizierung ausgeschlossen werden kann (Höhne 2010: 10).

---

<sup>4</sup> ISCO steht für „International Standard Classification of Occupations“.

<sup>5</sup> Einige Beispiele: Zum Stichtag 31.12.2012 hatte die Hallig Gröde, die kleinste Gemeinde in der Bundesrepublik Deutschland, 11 Einwohner. Das Saarland umfasst nur eine Großstadt, nämlich Saarbrücken. Und Bremen besteht nur aus zwei Gemeinden, Bremen und Bremerhaven.



Zentral bei der Bearbeitung der Daten sind vor allem zwei Merkmalskreise: zum einen der Ort der Datenerhebung und zum anderen personenbezogene Daten. Kenne ich die Gemeinde, in der eine Erhebung stattgefunden hat und weiß ich, dass eine Befragte z.B. spezielle Fachärztin ist (Zahnmedizin, Gynäkologie, Orthopädie), dann ist die Wahrscheinlichkeit groß, dass ich diese Person mit Hilfe von Zusatzwissen, z.B. aus den „Gelben Seiten“ im Internet, identifizieren kann. Die Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS; Beispiel 1 in Anhang 2) ist eine allgemeine Umfrage bei der detaillierte Informationen z.B. über die Berufstätigkeiten der Befragten erhoben werden (4-stelliger ISCO-Code). Würde man diese Daten mit detaillierten Regionalangaben weiterreichen, wäre das Reidentifizierungsrisiko für einige Befragungsteilnehmer hoch.

Wie die Beispiele in Anhang 2 zeigen, werden in der Regel die regionalen Angaben oder die Berufsangaben vergrößert. Das bedeutet, dass z.B. die Berufs-Codes nach den Vorgaben des Statistischen Bundesamtes von vier auf drei Stellen verkürzt werden, um lediglich größere Berufsgruppen abzubilden (Beispiel 2, Lebensverläufe). In vielen Fällen, wie dem des ALLBUS oder des Sozio-oekonomischen Panels (SOEP; Beispiel 3) werden die regionalen Angaben auf das Niveau des Bundeslandes vergrößert oder sogar Bundesländer zusammengelegt (im SOEP Rheinland-Pfalz und Saarland).

Bei Befragungen unter speziellen Populationen, wie HochschulabsolventInnen (Beispiel 4) oder WissenschaftlerInnen (Beispiel 5), kommen weitere Faktoren hinzu, die eine Anonymisierung aufwändiger machen. In den beiden genannten Fällen spielt der Rahmen der Datenerhebung eine wichtige Rolle. Absolvierten Befragte einen Prüfungsabschluss in einem „seltenen“ Fach<sup>6</sup>, wie Byzantinistik (7 Absolventen in 2013), Niederdeutsch (1 Absolventin in 2013) oder Blinden-/Sehbehindertenpädagogik (9 Absolventen in 2013), dann kann über die wahrscheinlich geringe Zahl an Hochschulen, die diese Fächer anbieten, zumindest ein regionaler Bezug zur Person hergestellt werden. Hier müssten die Fachbereiche vergrößert werden. Gleiches gilt für WissenschaftlerInnen z.B. an Universitäten. Diese nehmen insofern eine exponierte Stellung ein, da sie sich in der Regel mit vollständigen Lebensläufen auf eigenen Webseiten vorstellen.<sup>7</sup> Der Anhang 2 geht auf weitere Fälle und Maßnahmen ein.

### 2.2.2 Pseudonymisierung

Unter Pseudonymisierung versteht man nach BDSG (§3, Abs. 6a) „das Ersetzen des Namens und anderer Identifikationsmerkmale durch ein Kennzeichen zu dem Zweck, die Bestimmung des Betroffenen auszuschließen oder wesentlich zu erschweren.“ Dieses Ersetzen kann in Daten aus Studien mit quantitativen wie auch mit qualitativen Designs erfolgen. In quantitativen Datenbeständen erfolgt die Pseudonymisierung in der Regel durch das Ersetzen von Namen und Kontaktinformationen durch eine Befragtennummer (auch als formale Anonymisierung bezeichnet). In Transkripten offener Interviews, dem wahrscheinlich häufigsten Anwendungsfall in der qualitativen Sozialforschung, ist die Pseudonymisierung aufwändiger.

MitarbeiterInnen des Archivs für Lebenslaufforschung (ALLF), heute QualiService Bremen, haben ein Pseudonymisierungskonzept für archivierte Transkripte erstellt (siehe Anhang 2, Beispiel 9). Ziel ist es, dass „nicht nur die Namens- und Ortsangaben zu den befragten Personen verändert werden, sondern auch jegliche personenbezogenen Angaben von Dritten bzw. über Dritte, die zu einer Reidentifizierung der befragten Person führen könnten. Dazu gehörte auch das Verändern der Namen von Institutionen und das Unkenntlichmachen von Jahresangaben. Für besonders kritische Textpassagen oder sogar Einzelfälle sollte sich vorbehalten werden, diese eventuell komplett zu löschen.“<sup>8</sup>

<sup>6</sup> Zahlen aus Destatis (2013)

<sup>7</sup> Ähnlich herausgehobene Positionen nehmen auch andere Personen des öffentlichen Lebens z.B. Bundestagsabgeordnete, ManagerInnen großer Unternehmen, oder KünstlerInnen ein.

<sup>8</sup> Kurzdarstellung im Anhang zu Medjedovic, Witzel (2010)

Der in Bremen verwendete Pseudonymisierungsschlüssel ist mit einer Datenbank verbunden und erlaubt eine einheitliche Ersetzung von Informationen über verschiedene Interviews. Für Nutzer, der keinen Zugang zu dieser Datenbank haben, können die so bearbeiteten Transkripte als anonym angesehen werden. D.h. wesentlich für die Abschätzung eines Reidentifizierungsrisikos ist der Zugang zu den Informationen über die Pseudonyme. Sind diese nicht frei zugänglich und z.B. bei einem Datentreuhänder untergebracht, trägt dies zum Schutz der Daten bei.

### 2.2.3 Zugangskontrolle

Können Daten nicht in einer Weise anonymisiert oder pseudonymisiert werden, dass eine Weitergabe an Dritte für die wissenschaftliche Nutzung problemlos möglich ist, kann über einen eingeschränkten Zugang, eine sog. On-Site-Nutzung nachgedacht werden. Diese On-Site-Nutzung ist das Standardmodell z.B. bei Daten der amtlichen Statistik, die über die Forschungsdatenzentren (FDZ) des Bundes und der Länder angeboten werden.<sup>9</sup> Auch andere Datenhaltende Einrichtungen in Deutschland betreiben FDZ oder Datenservicezentren. Einen Überblick bietet der Rat für Sozial- und Wirtschaftsdaten (RatSWD).<sup>10</sup> Ein On-Site-Zugang bietet die Möglichkeit, Daten zur wissenschaftlichen Nutzung unter kontrollierten Bedingungen anzubieten. Dies bedeutet nicht, dass die Daten ohne Reduzierungen im Informationsgehalt auskommen. Allerdings bieten Nutzungsverträge und die Kontrolle der Analyseergebnisse durch betreuendes Personal einen Schutz vor der missbräuchlichen Verwendung von personenbeziehenden Daten. Personenbeziehbar meint hier, dass ein Bezug zu einer Befragungsperson durch Zusatzwissen eventuell möglich ist. Ein On-Site-Zugang stellt immer eine Abwägung zwischen einem Informationsverlust durch Anonymisierungsmaßnahmen und dem möglichst einfachen Zugang zu Forschungsergebnissen dar.

## 2.3 Datenschutz und Datenservice bei GESIS

Das Leibniz-Institut für Sozialwissenschaften GESIS ist eine Einrichtung, die sich unter anderem der Archivierung, also der Langfristsicherung, Dokumentation und dem Angebot sozialwissenschaftlicher Forschungsergebnisse widmet. Seit 2007 ist GESIS eine zentralisierte Einrichtung mit fünf Abteilungen, die den gesamten Lebenszyklus empirischer Forschung abdeckt. Das Institut ist eine sowohl Daten erhebende als auch Daten haltende Einrichtung, und bietet Dienstleistungen rund um den sog. Forschungszyklus an. Hierzu zählen Beratungen bei der Fragebogenerstellung, bei der Stichprobenziehung oder beim Gewichten von Daten, beim Pretesting (Test z.B. eines Fragebogens unter kontrollierten Bedingungen um etwa das Fragenverständnis zu untersuchen), bei Literaturrecherchen, sowie Schulungen zu verschiedenen methodischen und statistischen Themen.

GESIS bietet u.a. Umfragedaten, Sozialindikatoren, Daten der Historische Sozialforschung, ökonomische Zeitreihen, und Metadaten zur amtlichen Erhebungen (datenbeschreibende Informationen, die die Arbeit etwa mit den Mikrozensen erleichtern) für die Sekundäranalyse an. Hierbei kann, wie oben beschrieben, zwischen der Weitergabe von Daten an Dritte (Off-Site-Nutzung) und der Verwendung vor Ort im Secure Data Center (On-Site-Nutzung) unterschieden werden.

### 2.3.1 Weitergabe an Dritte (Off-Site-Nutzung)

Die Weitergabe an Dritte erfolgt in der Regel per Download nach einer Nutzerregistrierung. Angebote wie der Datenbestandskatalog bieten Übersichten und Datenzugänge. GESIS unterscheidet verschiede-

<sup>9</sup> Hinweise zum Angebot finden sich auf <http://www.forschungsdatenzentrum.de/> (Zugriff: 15.09.2014)

<sup>10</sup> <http://www.ratswd.de/>

ne Zugangsklassen, so dass in einigen Fällen eine formlose Genehmigung des Datengebers eingeholt oder vor Bereitstellung der Daten ein Datennutzungsvertrag unterzeichnet werden muss. Eine Übersicht über die Zugangsklassen und alle weiteren Informationen zur Archivierung bei GESIS bietet unsere Webseite: [www.gesis.org/datenarchivierung](http://www.gesis.org/datenarchivierung) (Zugriff: 15.09.2014).

### 2.3.2 Nutzung vor Ort (On-Site-Nutzung)

Für ein Angebot sensiblerer, quantitativer Daten vor Ort steht NutzerInnen das Secure Data Center (SDC) offen.<sup>11</sup> Das SDC wurde eingerichtet, um Datengebern die Möglichkeit zu eröffnen, ihre Daten in zwei Varianten zu archivieren. Es besteht nun die Option eine „Vollversion“ für die On-Site Nutzung sowie eine reduzierte Variante für den einfacheren Zugang über den Datenbestandskatalog anzubieten. Da Anonymisierungsmaßnahmen immer den Informationsgehalt von Daten reduzieren (s.o.), besteht nun größere Freiheit bei Entscheidungen über Art und Umfang dieser Maßnahmen. Es muss also nicht mehr nur ein eventuell auf ein Minimum reduzierter Datensatz angeboten werden.

---

<sup>11</sup> GML: <http://www.gesis.org/gml> (Zugriff: 15.09.2014); SDC: <http://www.gesis.org/sdc> (Zugriff: 15.09.2014)

### 3 Anhang 1: Übersicht Beispiele und kritische Informationen und mögliche Maßnahmen

<i>Nr</i>	<i>Beispiel</i>	<i>Typ</i>	<i>Klassifikation</i>	<i>Kritische Informationen</i>	<i>Maßnahmen</i>
1	ALLBUS	Querschnitt		Kleinräumige regionale Informationen, detaillierte Berufsangaben	Vergrößerung der regionalen Angaben auf der Ebene Bundesland
2	Lebensverläufe	Lebensverlauf, Querschnitt	In Ergänzung zum Querschnitt ALLBUS mehr Informationen zum Lebensverlauf	Kleinräumige regionale Informationen, detaillierte Berufsangaben, detaillierte Lebensverläufe	Zusätzlich zu regionalen Angaben Vergrößerung der klassifizierten Berufsangaben
3	SOEP	Lebensverlauf, fortlaufende Erhebung (Panel)	In Ergänzung zum Lebensverlauf fortlaufende Erhebung	Kleinräumige regionale Informationen, detaillierte Berufsangaben, detaillierte Lebensverläufe	Vergrößerung regionaler Angaben; Vertragsmodell für Nachnutzung
4	Absolventenbefragung an Hochschulen	Querschnitt	Spezielle Population	Spezieller Erhebungskontext durch Institution und Studienfach	Vergrößerung regionaler Angaben und der Informationen zu Studienfächern
5	Befragung Wissenschaftler-Innen	Querschnitt	Spezielle Population mit hoher Sichtbarkeit	Spezieller Erhebungskontext durch Institution, Fachgebiet und öffentliche Sichtbarkeit	Vergrößerung regionaler Angaben und der Informationen zu Fachbereichen
6	Schülerbefragung	Querschnitt	Spezielle Population mit besonderem rechtlichem Schutz	Spezieller Erhebungskontext durch Institution	Vergrößerung regionaler Angaben und der Informationen zu Schultypen
7	Kleinräumige Geographie		Prinzipielle Herausforderung aller Erhebungen	Merkmalskombination (Größe, Typ) kann zur Identifizierung einzelner Gemeinden führen.	
8	Offene Fragen		Mögliche Offenlegung einzelner genauer Informationen (z.B. Beruf)	Abhängig von Variable	Entfernen der offenen Angaben
9	Offene Interviews	Querschnitt oder Längsschnitt	In Ergänzung zu offenen Fragen mögliche Offenlegung z.B. des Lebensumfeldes	Abhängig von Befragungszielen	Z.B. Pseudonymisierung offener Angaben zu Namen, Orten, Einrichtungen etc.

## 4 Anhang 2: Fallbeispiele

---

Alle hier aufgeführten Fälle, bis aus das SOEP, betreffen Daten, die der Abteilung Datenarchiv von GESIS anvertraut wurden. Die meisten der genannten Fallbeispiele sind ohne Nennung des Namens der Studie dargestellt, da sich für sie getroffenen Maßnahmen auf andere, ähnliche Fälle übertragen lassen. Auch hier sei nochmals darauf hingewiesen, dass es keine Standardlösungen für die Erreichung eines datenschutzkonformen Informationsgehaltes von Forschungsergebnissen gibt. Jeder Fall bei dem es Zweifel gibt, muss einzeln geprüft werden.

### 4.1 Beispiel 1: Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS)

#### *Fallbeschreibung*

„Mit der Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) werden aktuelle Daten über Einstellungen, Verhaltensweisen und Sozialstruktur der Bevölkerung in der Bundesrepublik Deutschland erhoben. Seit 1980 wird alle zwei Jahre ein repräsentativer Querschnitt der Bevölkerung mit einem teils konstanten, teils variablen Fragenprogramm befragt. Die Daten stehen unmittelbar nach ihrer benutzergerechten Aufbereitung und Dokumentation allen Interessenten für Forschung und Lehre zur Verfügung (Webseite ALLBUS: <http://www.gesis.org/allbus>; Zugriff: 11.09.2014).“

Im ALLBUS werden alle zwei Jahre ca. 3.500 Personen über ihre individuelle soziale und wirtschaftliche Situation, die Situation ihres Haushaltes, Einstellungen und Meinungen, sowie im Anschluss an das International Social Survey Program (ISSP) zu ausgewählten Themen wie „Familie und Geschlechterrollen“ (2012), „Gesundheit“ (2011), oder „Umwelt“ (2010) befragt. Die Erhebung umfasst eine Vielzahl von Variablen zu den Befragten und ihren Haushalten, die in ihrer Gänze ein relativ detailliertes Bild der Lebensumstände bieten (vgl. die jeweils mit den Datensätzen publizierten Variablenreports, recherchierbar über <https://dbk.gesis.org/dbksearch/>; Zugriff: 11.09.2014). Ferner wird aus Qualitätsgründen großer Wert auf methodische Genauigkeit beim Sampling, der Befragung, und der Verarbeitung der Daten gelegt.

#### *Datenschutzproblem*

Drei Bereiche, zu denen im ALLBUS Daten erhoben werden – bzw. im Zuge der Feldarbeit entstehen – sind die berufliche Situation der Befragten (u.a. 4-stelliger ISCO-Code), die regionale Verortung der Befragten (Sampling Points, Wohnort), und Informationen rund um den eigentlichen Datenerhebungsvorgang (sog. Paradata, vgl. Kreuter, Casas-Cordero 2010). Diese Informationen sind teilweise nur dem Feldforschungsinstitut zugänglich (z.B. Adressen). Allerdings umfassen auch die so formal anonymisierten Daten noch ein Restrisiko der Re-Identifizierung für die Befragten. Außerdem wird beim ALLBUS großer Wert auf die informierte Einwilligung der Befragten gelegt, die z.B. den genauen Titel der Erhebung und die verantwortliche Forschungseinrichtung umfasst. Werden diese Informationen z.B. durch den Befragten an Dritte weitergegeben, die nicht am Forschungsprojekt beteiligt sind, erhalten diese Teilnahmekennntnis. Da der ALLBUS explizit für die breite Nutzung in Lehre und Forschung erhoben wird, spielt der Datenschutz eine große Rolle.

### *Mögliches Zusatzwissen*

Zusatzwissen, welches die Gefahr der De-Anonymisierung beim ALLBUS erhöht, kann aus verschiedenen Quellen stammen. So stellt das statistische Bundesamt eine Vielzahl von statistischen Informationen auch über kleine administrative Einheiten wie Gemeinden bereit. Mittlerweile haben viele, auch kleinste, Gemeinden eigene Seiten in Wikipedia. Eine Vielzahl von Portalen sammelt zudem im Internet verfügbare Angaben wie Adressen, Social-Media-Kontaktdaten, oder Berufsinformationen (besonders von all denjenigen, die aus Werbezwecken eigene Webseiten etc. betreiben). Eine genaue Beschreibung der räumlichen Umgebung der Befragten könnte durch Bilddienste wie Google Street View überprüft werden.

### *Lösung*

Die Daten des ALLBUS werden für die Herausgabe an ForscherInnen im Bereich der regionalen Informationen vergrößert. Mit dem Datensatz werden eine Variable zum Bundesland, die politische Gemeindegrößenklasse sowie die sog. BIK-Region (Klassifizierung des Wohnortes z.B. nach städtisch und ländlich) weitergegeben. Somit behält der ALLBUS sein inhaltliches und erhebungsmethodisches Analysepotential. Kleinräumigere Analysen, für die Angaben etwa zu Gemeinden oder Kreisen notwendig wären, sind auf Antrag on-site möglich. Geschützte Remote-Access-Lösungen sind in Vorbereitung.

## **4.2 Beispiel 2: Lebensverläufe**

### *Fallbeschreibung*

In einer Querschnitterhebung wurden ausführliche Lebensverläufe von Befragten einzelner Geburtsjahrgänge erhoben.

### *Datenschutzproblem*

Ähnlich wie beim SOEP werden in dieser Studie detaillierte Angaben zum Lebensverlauf gemacht: Herkunft, Ausbildung, beruflicher Werdegang, derzeitige Berufstätigkeit, häusliche Situation etc. Anders als beim fortgeschriebenen Panel handelt es sich hier nur um eine einmalige Befragung. Dennoch kann die Gesamtheit der Informationen ein Risiko für die Befragten darstellen. Ähnlich wie beim ALLBUS und dem SOEP handelt es sich um eine herausgehobene Studie, deren Titel den Befragten genau mitgeteilt wurde. Teilnahmekennntnis könnte also ein Problem sein, wenn die Befragten die Information über die Erhebung mit Dritten teilen. Wie in anderen Fällen spielen auch hier die Kombination detaillierter Berufs- und kleinräumiger Regionalangaben eine wichtige Rolle.

### *Mögliches Zusatzwissen*

Informationen über die Verteilung spezieller Berufe (z.B. Handwerker) in kleinen Gemeinden lassen sich über frei zugängliche Quellen besonders im Internet nachvollziehen (u.a. „Gelbe Seiten“).

### *Lösung*

Die kleinste im Datensatz belassene Regionalvariable ist das Bundesland. Zudem werden die Berufsangaben durch Codes ersetzt. Nach der Klassifikation des Statistischen Bundesamtes (1992) wurden Be-

rufscodes nur dreistellig im Datensatz belassen (z.B. „853 Krankenschwestern/-pfleger, Hebammen/Entbindungspfleger“ statt detaillierterer Aufschlüsselung).<sup>12</sup>

### 4.3 Beispiel 3: Fortlaufende Erhebung mit identischen Personen (SOEP)

#### *Fallbeschreibung*

Das SOEP ist ein zentraler Bestandteil der Sozialberichterstattung in der Bundesrepublik Deutschland. Gestartet 1984 im Rahmen des Sonderforschungsbereichs 3 am Deutschen Institut für Wirtschaftsforschung, umfasste es nach den ersten drei Wellen rund 220 Variablen. Sollte es ursprünglich ca. 5.000 Haushalte umfassen (Hanefeld 1987:141f.), so repräsentierte es 2013 jährlich etwa 30.000 Befragte in fast 11.000 Haushalten.<sup>13</sup> Bereits in den ersten drei Wellen wurde das ursprüngliche Fragenprogramm um Items etwa zum Wohnumfeld, beruflicher Veränderung, oder sozialer Herkunft erweitert (siehe Fragenprogramm bei Hanefeld 1987:272-280). In den 30 Jahren seiner Existenz wurden zahlreiche zusätzliche Querschnittserhebungen durchgeführt. Ferner werden im sog. SOEPlong, dem longitudinalen Bestand, jährlich über 2.000 Variablen um aktuelle Daten erweitert (Gerstorff, Schupp (Eds.) 2013). Analysefiles können mit Hilfe des Informationsportals SOEPinfo erstellt werden.

#### *Datenschutzproblem*

Das SOEP bietet einen sehr umfassenden Einblick in individuelle Entwicklungen und die Veränderungen von Haushalten. Durch zahlreiche Zusatzmodule übersteigt der Datenumfang das üblich Maß von Querschnitten wie dem ALLBUS um ein Vielfaches. V.a. die individuellen Lebensverläufe, die teilweise über Jahrzehnte nachvollzogen werden, stellen eine große Herausforderung dar.

#### *Mögliches Zusatzwissen*

Ähnlich wie beim ALLBUS können auch hier eine Vielzahl von möglichen Informationsquellen in Betracht gezogen werden.

#### *Lösung*

Das SOEP wird in drei abgestuften Varianten mit einem Datennutzungsvertrag und einer Datenschutzvereinbarung angeboten. Es wird ein Scientific Use File (SUF) für die Off-Site-Nutzung der Daten erzeugt, der keine kleinräumigen Regionalangaben enthält. Gemeindegrößenklassen und Raumordnungsregionen können nur mit einer Zusatzvereinbarung genutzt werden. Ein Angebot von z.B. Kreisangaben findet nur On-Site statt (Frick et al. 2010).

### 4.4 Beispiel 4: Absolventenbefragung an Hochschulen

#### *Fallbeschreibung*

In einem Wintersemester wurden an fast 50 Hochschulen in 12 von 16 deutschen Bundesländern Absolventenbefragungen durchgeführt. Von den bundesweit ca. 309.000 AbsolventInnen des entspre-

---

<sup>12</sup> Klassifikationen bei Destatis unter: <https://www.destatis.de/DE/Methoden/Klassifikationen/Klassifikationen.html> (Zugriff: 29.10.2014)

<sup>13</sup> SOEP: [http://www.diw.de/de/diw\\_02.c.222847.de/handbuch\\_ueberblick.html](http://www.diw.de/de/diw_02.c.222847.de/handbuch_ueberblick.html) (Zugriff: 29.10.2014)

chenden Jahrgangs schlossen etwa 104.000 (35 Prozent) ihr Studium an den beteiligten Hochschulen ab. Es wurde eine Vollerhebung dieser Absolventen angestrebt, an der sich jedoch nicht alle Fachbereiche und Abschlussarten beteiligten. Somit kamen ca. 82.000 AbsolventInnen für eine Erhebung in Frage. Die Netto-Stichprobe bildeten dann ungefähr 78.000 Absolventen.

Der Fragebogen bestand aus Kernfragen zu den Hochschulen, aus einem Pool optionaler Fragen und eigener Fragen. Zentrale Themen der Befragung waren u.a. der Verbleib nach Studienabschluss, die Beschäftigungssuche, sowie die berufliche Tätigkeit und Kompetenzen.

Die AbsolventInnen wurden von den Hochschulen auf dem Postweg kontaktiert. Die Teilnahme erfolgte freiwillig und auf den Webseiten der beteiligten Hochschulen wurden AnsprechpartnerInnen und Datenschutzhinweise veröffentlicht. Befragt wurde mit Papier- und Online-Fragebögen.

Für die weitere Verarbeitung der Daten gab es eine rechtliche Einschränkung, da sich das Forschungsprojekt vertraglich verpflichtete, die Daten nur so zu verarbeiten, dass Rückschlüsse auf die Kooperationshochschulen nicht möglich sind. Ferner handelt es sich bei den ausgewählten Hochschulen nicht um eine Zufallsstichprobe. Alle Hochschulen sind in der Dokumentation der Studie aufgelistet.

### *Datenschutzproblem*

In einem ähnlich gelagerten Fall geht die Hochschul-Informations-System GmbH (2006:13) davon aus, dass der „wesentliche Risikofaktor für die faktische Anonymität von Absolventendaten [...] in der gleichzeitigen regionalen und sachlichen Tiefengliederung“ besteht. D.h. praktische Maßnahmen sollten sich auf diese Kombination beziehen. Kenngrößen, denen bei einer Risikoanalyse besondere Aufmerksamkeit zukommen sollte, sind einzigartige Ausprägungen in der Grundgesamtheit:

- Gibt es Hochschulen, die aufgrund ihrer Ausrichtung oder ihres regionalen Kontexts einmalig sind?
- Gibt es einzigartige Abschlüsse, die nur an einer Hochschule vergeben werden?
- Gibt es seltene Abschlüsse, die nur wenigen Hochschulen zugeordnet werden können? Durch Ausschluss der nicht am Sample beteiligten Hochschulen könnten einzelne Institutionen identifizierbar sein.

### *Mögliches Zusatzwissen*

Als Zusatzwissen könnten frei verfügbare Informationen und Daten des Statistischen Bundesamtes herangezogen werden (z.B. Daten zu abgelegten und bestandenen Prüfungen an deutschen Hochschulen im Prüfungsjahr, Statistisches Bundesamt). Viele Hochschulen bieten ferner detaillierte Übersichten zu Studienabschlüssen auf ihren Webseiten an. Eine Kombination von Informationen über die Hochschule, an der ein Abschluss erworben wurde, die Abschlussart und die genaue Fächerbezeichnung scheint lediglich die Prüfungsstatistik zu bieten. Diese ist jedoch ausschließlich in Forschungsdatenzentren der Statistischen Ämter nutzbar und scheidet so als Datenbestand für einen Abgleich mit der Absolventenbefragung aus.

Das frei verfügbare Zusatzwissen lässt sich jedoch nicht systematisch für Reidentifizierungsversuche nutzen, da es lediglich die Möglichkeit bietet, nicht häufig absolvierte Studienfächer an kleinen Hochschulen zu bestimmen.

### *Lösung*

- Aufgrund vertraglicher Vereinbarungen zwischen dem Projekt und den kooperierenden Hochschulen wurden Bundesländer teilweise zusammengefasst. Hochschulspezifische Angaben (z.B. nicht angebotene Fächer) wurden über für alle Fälle zusammengefasst. Hiermit soll die Identifizierung einzelner Hochschulen verhindert werden.
- Angaben der Studienfächer wurden nach Vorgaben des Statistischen Bundesamtes vercodet und gruppiert. Kleine Fallzahlen bei bestimmten Studienfachrichtungen konnten vergrößert werden (Bsp. Zusammenfassung seltener Studienfächer wie Baltistik, Meteorologie oder Landespflege). Solche Vergrößerungen können mit Hilfe von Angaben des Statistischen Bundesamtes (z.B. „Prüfungen an Hochschulen“) erfolgen.



- Offene Angaben zu Stipendien wurden gelöscht.
- Eine regionale Zuordnung war im Datensatz über die Angabe des zulassenden Kreises bzw. der kreisfreien Stadt durch Nennung des Kfz-Kennzeichens (z.B. K für Köln) gegeben. Diese Angaben wurden aus dem Datensatz entfernt.
- Berufe wurden teilweise im Klartext erfasst. Diese wurden in ISCO-Codes überführt.

## 4.5 Beispiel 5: Elitenbefragungen

### *Fallbeschreibung*

In einer Querschnittserhebung wurden WissenschaftlerInnen an Hochschulen und hochschulnahen Forschungseinrichtungen zu ihren Karrieren, zur Forschungssituation, zu wissenschaftlichem Fehlverhalten und zur Forschungsförderung befragt.

### *Datenschutzproblem*

Bei WissenschaftlerInnen an Hochschulen und hochschulnahen Forschungseinrichtungen handelt es sich um exponierte Personen des öffentlichen Lebens. Häufig werden sie z.B. als Experten von der Politik und den Medien für Expertisen herangezogen. Teilweise gibt es nur eine Universität in einem Bundesland (Saarland), spezielle Hochschulen nur an einem Standort (z.B. Deutsche Sporthochschule Köln), oder einzelne Fächer nur an wenigen Hochschulen. Durch ihre Spezialisierung in Fachdisziplinen sind WissenschaftlerInnen damit ggf. über die Institutionen reidentifizierbar. Wenn die Einrichtung bestimmt werden kann und Geschlecht, Alter und Forschungsrichtung bekannt sind, ist eine Bestimmung der Person relativ einfach möglich. In einigen Fällen können auch Nationalitäten oder Staatsangehörigkeiten ein Problem darstellen.

### *Mögliches Zusatzwissen*

Lebensläufe von ForscherInnen sind üblicherweise über die Webseiten von Hochschulen und Forschungseinrichtungen verfügbar. Forschungsförderer stellen mittlerweile Datenbanken mit den Namen von EmpfängerInnen von Drittmitteln bereit (z.B. die GEPRIS Datenbank der DFG).<sup>14</sup> Das Statistische Bundesamt bietet eine Vielzahl von Informationen zu Studiengängen an öffentlichen Hochschulen. (Siehe auch Beispiel 4 oben für weitere Informationen zu Zusatzwissen.)

### *Lösung*

- Für das Datenangebot wurden zwei Varianten gewählt. Eine reduzierte Version umfasst keine kleinräumigen geografischen Einheiten, keine detaillierten Informationen zur Forschungsförderung und nur vergrößerte Informationen zum Fachbereich. Eine detailliertere Variante wird nur zur Nutzung vor Ort angeboten.
- Für beide Varianten ist jeweils der Abschluss eines Datennutzungsvertrages nötig.

<sup>14</sup> <http://gepris.dfg.de/gepris/OCTOPUS?context=projekt&index=0&task=showSearchSimple> (Zugriff: 11.09.2014)

## 4.6 Beispiel 6: Schülerbefragung

### *Fallbeschreibung*

Für eine Befragung von SchülerInnen, Eltern und LehrerInnen in der Bundesrepublik wurde ein dreistufiges Auswahlverfahren angewandt. Die erste Stufe betrifft die Schulen, die zweite Stufe die Klassen, die dritte die SchülerInnen. Die SchülerInnen sollten 14 Jahre alt sein.

Für die Auswahl der Schulen wurden zunächst komplette Übersichten aller Einrichtungen pro Bundesland angelegt. In jeder Schule wurde eine Ansprechperson ausgewählt. Beim Zugang zu SchülerInnen gelten einschränkende gesetzliche Regelungen, in aller Regel die Schulgesetze. So sieht etwa das Schulgesetz NRW<sup>15</sup> (§120, Abs.1) vor, dass gespeicherte personenbezogene Daten in der Schule nur den Personen zugänglich gemacht werden dürfen, „die sie für die Erfüllung ihrer Aufgaben benötigen.“ Befragungen von SchülerInnen dürfen nur zu Zwecken der sonderpädagogischen Förderung oder für Maßnahmen zur Qualitätsentwicklung und Qualitätssicherung durchgeführt werden (§120, Abs.3). Zur wissenschaftlichen Forschung heißt es (§120, Abs.4):

„Andere wissenschaftliche Untersuchungen, Tests und Befragungen sind nur mit Einwilligung im Rahmen des Absatz 2 Sätze 2 und 3 zulässig, wenn dadurch die Bildungs- und Erziehungsarbeit und schutzwürdige Belange einzelner Personen nicht beeinträchtigt werden oder die Anonymität der Betroffenen gewahrt bleibt. Die Entscheidung trifft die Schulleiterin oder der Schulleiter. In Angelegenheiten besonderer oder überörtlicher Bedeutung ist die obere Schulaufsichtsbehörde zu unterrichten.“

Die Schulen wurden nach der Auswahl schriftlich über das Vorhaben informiert und in den teilnehmenden Einrichtungen war es in der Regel möglich, Klassen zufällig auszuwählen. Es wurde in allen Klassen versucht, Totalerhebungen durchzuführen. Für eine Befragung der SchülerInnen mussten die Eltern eine Einverständniserklärung unterschreiben. Zusätzlich wurde versucht, die Eltern der teilnehmenden SchülerInnen und die LehrerInnen der Kinder zu interviewen.

### *Datenschutzproblem*

Das Problem der möglichen Reidentifizierung von SchülerInnen betrifft zwei Variablen: (a) die Region und (b) die Schulen. Hat eine Schule in einer größeren administrativen Region eindeutige Merkmale, könnte sie trotz der Vergrößerung der regionalen Merkmale eventuell bestimmt werden. Im Sample sind Schultypen erfasst. So wurden etwa sog. Rudolf-Steiner-Schulen angesprochen (Synonym für Waldorfschulen). In Deutschland gibt es hiervon 55.<sup>16</sup> (Selbst in den bevölkerungsmäßig kleinen Bundesländern wie Bremen und dem Saarland gibt es jedoch mehr als eine solche Schule.)

### *Mögliches Zusatzwissen*

Um beim Beispiel NRW zu bleiben, so bietet das Landschulministerium eine Webseite mit Auswahlmöglichkeiten von Schulen z.B. nach Typen an.<sup>17</sup> Wie das Beispiel Waldorfschulen zeigt, gibt es für spezielle Schultypen eigene Webseiten.

---

<sup>15</sup> Schulgesetz NRW (SchulG), Stand vom 01.09.2014:  
<http://www.schulministerium.nrw.de/docs/Recht/Schulrecht/Schulgesetz/Schulgesetz.pdf> (Zugriff: 15.09.2014)

<sup>16</sup> Übersicht der Waldorfschulen: [www.waldorfschule.de](http://www.waldorfschule.de) (Zugriff: 19.09.2014)

<sup>17</sup> Schulsuche beim Schulministerium NRW: <https://www.schulministerium.nrw.de/BP/SchuleSuchen> (Zugriff: 19.09.2014)

### *Lösung*

Für den Datenzugang wurde als kleinste administrative Einheit das Bundesland gewählt. Für eine Variante der Daten zur Nutzung durch Dritte auf eigenen Rechnern wurden die Schultypen vergrößert. Ferner wurden aus anderen Gründen Angaben zur Herkunft der Eltern in dieser Variante der Daten unterdrückt.

## 4.7 Beispiel 7: Kleinräumige Geographie

### *Fallbeschreibung*

In einer bundesweit durchgeführten Studie zur Bundestagswahl 2009 wurden auch Wahlkreiskennziffern erhoben. Durch die Kombination dieser Kennziffer mit der sog. Gemeindegrößenklasse konnten eventuell einzelne kleine Gemeinden bis 2000 oder bis 5000 Einwohner identifiziert werden. Hierdurch konnte sich, in Kombination mit genauen Berufsangaben oder anderen ein-eindeutigen Informationen ein Identifizierungsrisiko ergeben. Generell wird davon ausgegangen, dass eine kleinräumige geographische Angabe in Kombination mit anderen Merkmalen eventuell das Identifizierungsrisiko einer befragten Person erhöht.

### *Datenschutzproblem*

Bei der Bundestagswahl 2009 war Deutschland in 299 Wahlkreise aufgeteilt, die 12.227 Gemeinden umfassten.<sup>18</sup> Es ergaben sich hier zwei "Problemfälle"<sup>19</sup>:

- 1) In Hessen existierten zum Zeitpunkt der Bundestagswahl 11 Gemeinden mit weniger als 2000 Einwohnern.
- 2) In NRW gab es 3 Gemeinden mit 2000-5000 Einwohnern. (Gemeinden mit weniger als 2000 Einwohnern existieren in NRW nicht.)

### *Mögliches Zusatzwissen*

Das Statistische Bundesamt und frei zugängliche Plattformen wie Wikipedia bieten eine Vielzahl von Informationen über die politische Struktur des Bundesgebietes. Eine Vielzahl auch kleiner Gemeinden pflegen z.B. eigene Wikipedia-Webseiten.

### *Lösung*

Da die Daten einer großen Zahl von wissenschaftlichen NutzerInnen zur Verfügung gestellt werden sollten, wurde vorgeschlagen, die beiden kleinsten Gemeindegrößenklassen (bis 2.000 sowie 2.000 bis 5.000) mit der nächst größeren (5.000 bis 20.000) zusammenzufassen.

---

<sup>18</sup> Strukturdaten für die Wahlkreise, in: Der Bundeswahlleiter (Hg.), Wahl zum 17. Deutschen Bundestag am 27. September 2009. Heft 1, Wiesbaden 2009, S. 208

<sup>19</sup> Gemeinden und Bevölkerung am 31.12.2007, in: Statistisches Bundesamt (Hg.), Statistisches Jahrbuch 2009, Wiesbaden 2009, S. 40f.

## 4.8 Beispiel 8: Offene Fragen

### *Fallbeschreibung*

In einer nationalen Wahlstudie eines europäischen Landes wurden bei den Befragten offene Berufsangaben erhoben und dem Datensatz numerische Codes für die Gemeinden, in denen befragt wurde, beigelegt.

### *Datenschutzproblem*

In den offenen Berufsangaben fand sich u.a. die Antwort „Vize-Bürgermeister“. In Kombination mit der Gemeindekennziffer hätte die befragte Person eindeutig identifiziert werden können.

### *(Mögliches) Zusatzwissen*

Informationen über die Identität von BürgermeisterInnen finden sich auf den Webseiten von Gemeinden im Internet.

### *Lösung*

Die offenen Berufsangaben wurden aus dem Datensatz entfernt und die geographischen Angaben auf ein nächst höheres Niveau (vergleichbar dem deutschen Regierungsbezirk, NUTS 2 Region) vergrößert.

## 4.9 Beispiel 9: Offene Interviews (qualitative Forschung)

### *Fallbeschreibung*

Bei diesem Beispiel handelt es sich um kein einzelnes Vorhaben, sondern um eine generelle kurze Darstellung der Probleme bei der qualitativen Forschung mit offenen Interviewformen.

### *Datenschutzproblem*

Offene Interviews werden oft mit wenigen Personen und in wenig standardisierter Form durchgeführt. Dies ist der Tatsache geschuldet, dass die qualitative Perspektive sich u.a. stärker der sozialen Konstruktion von Wirklichkeit widmet (Flick et al. 2010a; zu Auswahlverfahren besonders Merkmals 2010). Die qualitative Forschung ist „trotz der zunehmenden Bedeutung visueller Datenquellen wie Foto oder Film überwiegend eine Textwissenschaft“, die „für die Mehrzahl ihrer (hermeneutischen) Interpretationsverfahren auf das Medium des Textes als Arbeitsgrundlage angewiesen (Flick et al. 2010:24)“ ist.<sup>20</sup> Neben den Transkripten offener Interviews gibt es häufig zusätzlich Texte, die den Forschungsablauf beschreiben. Man spricht u.a. von Feld- oder Methodenberichten.

In offenen Interviews und den zugehörigen Feldnotizen oder Feldberichten werden nach Einschätzung vieler ForscherInnen detaillierte Informationen der Befragten erhoben (Ergebnis u.a. einer Machbarkeitsstudie, Opitz, Mauer 2005; Hopf (2010) geht auf Fragen der Forschungsethik in qualitativer Forschung ein). Befragte haben im Interview die Möglichkeit beliebig detaillierte Angaben zu Biografie und Erlebnissen zu machen.

---

<sup>20</sup> Zu den verschiedenen Methoden der qualitativen Forschung, siehe Flick et al. (2010b), S. 332–587.

*Mögliches Zusatzwissen*

Da oftmals Personen aus kleinen administrativ-geografischen Einheiten (z.B. Gemeinden) und engen sozialen Kontexten (z.B. SchülerInnen einer Schule, BesucherInnen einer Institution, NutzerInnen spezieller klinischer Einrichtungen) befragt werden, gilt hier in erhöhtem Maße das Reidentifizierungsrisiko über den Zugang zu Informationen über diesen Erhebungskontext.

*Lösung*

Watteler (2010) bietet eine Einführung in datenschutzrechtliche Fragen und praktische Maßnahmen der Anonymisierung qualitativer Forschungsergebnisse. Liebig et al. (2014) widmen sich besonders den rechtlichen Aspekten und hier speziell dem Einverständnis der Befragten. Das von Andreas Witzel aufgebaute Archiv für Lebenslaufforschung an der Universität Bremen hat seinen Fokus erweitert und bietet heute als QualiService auch praktische Anleitungen für die Anonymisierung von Transkripten (Kretzer 2013).

## 5 Anhang 3: Liste möglicher Identifizierungsmerkmale und Regionalangaben

---

### 5.1 Direkte Identifizierungsmerkmale

---

- a) Namen
  - b) Anschrift
  - c) Telefonnummer
  - d) Kfz-Kennzeichen (*Mögliche Identifizierung Fahrzeughalter*)
  - e) Personalausweisnummer
  - f) Sozialversicherungsnummer
  - g) E-Mail Adresse (*z.B. berufliche mit Vor- und/oder Nachnamen*)
  - h) Feste IP-Adresse
  - i) Ein-eindeutige Berufsbezeichnung (*z.B. Präsidentin, Rektorin, Direktorin etc. in Kombination mit Name Arbeitgeber*)
- 

### 5.2 Indirekte Identifizierungsmerkmale

---

- j) Offene Berufsangabe  
(*ggf. Name Unternehmen oder eindeutige Berufsangabe*)
  - k) Offene Angabe zu Schul- und Berufsbildung  
(*ggf. Name der Bildungseinrichtung*)
  - l) Karriereangaben im Lebensverlauf
  - m) Geburtsland
  - n) Staatsangehörigkeit
  - o) Muttersprache
- 

### 5.3 Regionalangaben

---

- p) Postleitzahl
  - q) Kreiskennziffer
  - r) Gemeindekennziffer
  - s) Ortsnamen (*auch Ortsnamen in der Dokumentation beachten*)
  - t) Stadtteilnamen
  - u) Bundesland oder Regierungsbezirk in Kombination mit Gemeindegrößenklasse oder Boustedt-Regionen
-

## 6 Literatur

---

- Bethlehem (2009): Bethlehem, Jelke, Applied Survey Methods. A Statistical Perspective, Hoboken 2009
- Destatis (2013): Statistisches Bundesamt, Prüfungen an Hochschulen. Fachserie 11 Reihe 4.2, Wiesbaden  
[https://www.destatis.de/DE/Publikationen/Thematisch/BildungForschungKultur/Hochschulen/PruefungenHochschulen2110420137004.pdf?\\_\\_blob=publicationFile](https://www.destatis.de/DE/Publikationen/Thematisch/BildungForschungKultur/Hochschulen/PruefungenHochschulen2110420137004.pdf?__blob=publicationFile) (Zugriff: 15.09.2014)
- Flick et al. (2010a): Flick, Uwe; Ernst von Kardorff; Ines Steinke, Was ist qualitative Forschung? Einleitung und Überblick, in: Flick et al. (2010b), S. 13–29
- Flick et al. (2010b): Flick, Uwe; Ernst von Kardorff; Ines Steinke, Qualitative Forschung. Ein Handbuch, 8.Aufl., Hamburg 2010
- Frick et al. (2010): Frick, Joachim R.; Jan Goebel; Hansjörg Haas; Peter Krause; Ingo Sieber; Michaela Engelmann, Verfahren für den Datenschutz beim Zugang zu den SOEP-Daten innerhalb und außerhalb des DIW Berlin, DIW Berlin  
[http://www.diw.de/documents/dokumentenarchiv/17/diw\\_01.c.347090.de/soep\\_datenschutzverfahren.pdf](http://www.diw.de/documents/dokumentenarchiv/17/diw_01.c.347090.de/soep_datenschutzverfahren.pdf) (Zugriff: 11.09.2014)
- Graumann (2006): Graumann, Sigrid, Forschungsethik, in: Düwell, Marcus; Christoph Hübenthal; Michael H. Werner (Hrsg.), Handbuch Ethik, 2.Aufl., Stuttgart; Weimar, S. 253–258
- HIS 2006: Fabian, Gregor; Karl-Hein Minks, Dokumentation des Scientific Use Files „HIS-Absolventenpanel 1997“, Hannover 2006 (Hochschul-Informationssystem (HIS), Hannover: HIS-Absolventenpanel 1997. GESIS Datenarchiv, Köln, Deutschland, ZA4863 Dokumentation)
- Gerstorff, Schupp (Eds.) (2013): Gerstorff, Sandra; Jürgen Schupp (Eds.), SOEP Wave Report 2013, DIW Berlin [http://www.diw.de/documents/publikationen/73/diw\\_01.c.470126.de/wave\\_report\\_2013.pdf](http://www.diw.de/documents/publikationen/73/diw_01.c.470126.de/wave_report_2013.pdf) (Zugriff: 11.09.2014)
- Höhne (2010): Höhne Jörg, Verfahren zur Anonymisierung von Einzeldaten, Wiesbaden (= Statistisches BA (Hrsg.), Statistik und Wirtschaft, Bd.16)  
[https://www.destatis.de/DE/Publikationen/StatistikWissenschaft/Band16\\_AnonymisierungEinzeldaten\\_1030816109004.pdf?\\_\\_blob=publicationFile](https://www.destatis.de/DE/Publikationen/StatistikWissenschaft/Band16_AnonymisierungEinzeldaten_1030816109004.pdf?__blob=publicationFile) (Zugriff: 15.09.2014)
- Hopf (2010): Hopf, Christel, Forschungsethik und qualitative Forschung, in: Flick et al. (2010b), S. 589–600
- Kretzer (2013): Kretzer, Susanne, Arbeitspapier zur Konzeptentwicklung der Anonymisierung/Pseudonymisierung in Qualiservice, Bremen  
[http://www.qualiservice.org/fileadmin/templates/qualiservice/Anonymisierungskonzept\\_Arbeitspapier.pdf](http://www.qualiservice.org/fileadmin/templates/qualiservice/Anonymisierungskonzept_Arbeitspapier.pdf) (Zugriff: 15.09.2014)
- Kreuter, Casas-Cordero (2010): Kreuter, Frauke; Carolina Casas-Cordero, Paradata, Berlin (= RatSWD Working Paper, No.136)  
[http://www.ratswd.de/download/RatSWD\\_WP\\_2010/RatSWD\\_WP\\_136.pdf](http://www.ratswd.de/download/RatSWD_WP_2010/RatSWD_WP_136.pdf) (Zugriff: 11.09.2014)
- Liebig et al. (2014): Liebig, Stefan; Tobias Gebel; Matthias Grenzer; Julia Kreusch; Heidi Schuster; Ralf Tschewinka; Oliver Watteler; Andreas Witzel, Datenschutzrechtliche Anforderungen bei der Generierung und Archivierung qualitativer Interviewdaten. Erarbeitet und verfasst von der Arbeitsgruppe Datenschutz und qualitative Sozialforschung, Berlin (= RatSWD Working Paper, No.238)  
[http://www.ratswd.de/dl/RatSWD\\_WP\\_238.pdf](http://www.ratswd.de/dl/RatSWD_WP_238.pdf) (Zugriff: 15.09.2014)

- Medjedovic, Witzel (2010): Medjedovic, Irena; Andreas Witzel, Wiederverwendung qualitativer Daten. Archivierung und Sekundärnutzung qualitativer Interviewtranskripte, Wiesbaden. Download Anhang zum Anonymisierungskonzept: <http://link.springer.com/book/10.1007/978-3-531-92403-8> (Zugriff: 15.09.2014)
- Merkens (2010): Merkens, Hans, Auswahlverfahren, Sampling, Fallkonstruktion, in: Flick et al. (2010b), S. 286-299
- Metschke, Wellbrock (2003): Metschke, Rainer; Rita Wellbrock, Datenschutz in Wissenschaft und Forschung, 3. Aufl., Berlin (Berliner Beauftragter für Datenschutz und Informationsfreiheit, Hrsg.) [www.datenschutz-berlin.de/attachments/47/Materialien28.pdf](http://www.datenschutz-berlin.de/attachments/47/Materialien28.pdf) (Zugriff: 11.02.2014)
- Opitz, Mauer (2005): Opitz, Diane; Reiner Mauer, Erfahrungen mit der Sekundärnutzung von qualitativem Datenmaterial – Erste Ergebnisse einer schriftlichen Befragung im Rahmen der Machbarkeitsstudie zur Archivierung und Sekundärnutzung qualitativer Interviewdaten, in: FQS Vol.6(1) 2005 <http://www.qualitative-research.net/index.php/fqs/rt/prINTERfriendly/510/1100> (Zugriff: 15.09.2014)
- Palandt (2008): Palandt, Otto; et al., Bürgerliches Gesetzbuch: mit Einführungsgesetz (Auszug) und andere, 67., neu bearb. Aufl., München 2008
- Terwey, Baltzer (2012): Terwey, Michael; Stefan Baltzer, ALLBUS 2012 – Variable Report, GESIS Köln <https://dbk.gesis.org/dbksearch/download.asp?db=D&id=49502> (Zugriff: 11.09.2014)
- Watteler (2010): Watteler, Oliver, Datenschutz und die Archivierung von Daten der qualitativen empirischen Sozialforschung, Kp. 3.2 in: Medjedovic, Irena; Andreas Witzel, Wiederverwendung qualitativer Daten. Archivierung und Sekundärnutzung qualitativer Interviewtranskripte, Wiesbaden, S. 63-85
- Wirth (1992): Wirth, Heike, Die faktische Anonymität von Mikrodaten: Ergebnisse und Konsequenzen eines Forschungsprojektes, in: ZUMA-Nachrichten (30) 1992, S. 7-44
- Wirth (2003): Wirth, Heike, Szenarien für Angriffe auf wirtschaftsstatistische Einzeldaten – Ein Überblick, in: Ronning, Gerd; Roland Gnos, Anonymisierung wirtschaftsstatistischer Einzeldaten, Wiesbaden 2003, S. 11-24